Reply to Referee # 1

"The manuscript analyses the role of bias correction in ENSEMBLES regional scenarios on the temperature response. Contrary to the so-called delta-method, quantile mapping modifies the mean model response. With an original linear approach, the authors show that the new response is more reliable than the un-corrected model response. As quantile mapping (or similar methods) is a "necessary evil" for driving impact models, this study is a major contribution to the climate impact community. The presentation is clear, with relevant citations. I recommend the manuscript for publication with minor corrections:"

Thank you for your very encouraging and constructive comments! We added our responses below your original comments:

"1. page 6 line 23 (and also title): it is clear that the approach can be extended to daily min and max temperature. But the application to precipitation is not as straightforward as the authors claim. Indeed the model error is generally: too many drizzle days and underestimated heavy precipitation. The notion of "error slope" is not adapted. Perhaps precipitation can be replaced by its logarithm or another function, but I hardly see a linear approach as in the present study. In addition, some models at some locations produce less rain days than in the observation, making quantile mapping not applicable (but applicable with a probabilistic approach). Precipitation correction is very important for impact studies (even more than temperature correction in many applications). Indeed, the sign of the response may be reversed after correction, because both the sign of the error and the sign of the response may change from low to high precipitation. I suggest to specify in the title that this study is devoted to temperature, to state in the perspectives that this approach could be extended to other variables, and I encourage the authors to prepare a second paper on precipitation correction."

We agree that the application of QM to precipitation has several specific issues and that our results for temperature cannot automatically be assumed to be valid for precipitation as well. To avoid such misinterpretation, we changed the text as suggested by the referee:
a) The title was adopted and reads now: The effect of empirical-statistical correction of intensity-dependent model errors on the temperature climate change signal.
b) Similar modifications were applied to the first sentence of the abstract and the last paragraph of the introduction.
c) On page 6 (section 2.2), the relevant sentences read now: In the following, we show the results for daily mean temperature, but the analysis of daily minimum and maximum temperatures gives very similar results. The application of our analysis to other parameters like, e.g., precipitation is basically straight forward, but the linearization applied in section 4 can be expected to be less appropriate for precipitation than for temperature. Further investigation is needed to fully reveal the effect of QM on the precipitation CCS. The major motivation for focusing on temperature here is its relatively simple error characteristic and its significant climate trend, which facilitates the demonstration of the effect of QM on the CCS.

"2. page 8 line 15: noisy tails (a funny typo)"

Thanks, corrected.

"3. page 17, line 9: you can mention that the new centennial reanalyses (NOAA and ECMWF) offer a good test bed for this time-invariance"

We added this information on page 6 (section 2.1), where the assumption of time-invariant model errors is discussed the first time. The respective sentences read now:

However, in a strict interpretation, the results and conclusions of this study are only valid under the assumption of time-invariant model errors and it is still issue to further investigation to determine the severity of this restriction. Although such investigation is outside the scope of our study, we want to mention that the new centennial re-analyses of ECMWF (ERA-20C) and NOAA-CIRES (V2c) offer a promising new test-bed for the investigation of the long-term stability of model error characteristics.

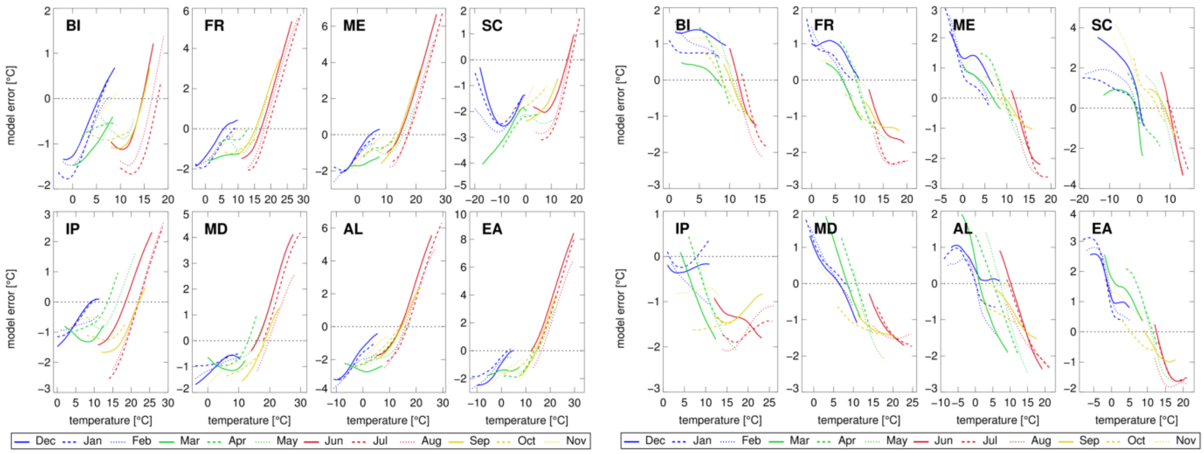
Reply to Referee # 2

“This study discusses the effect of empirical-statistical bias correction methods (quantile mapping, QM) on the change signals of climate simulations; in fact it has been previously shown that bias correction can alter the mean temperature climate change signal derived from multi-model ensembles in Europe. By means of an analytical analysis of the model error and its dependence on the value of simulated variable, the authors claim that the climate signal is artificially inflated by intensity-dependent model errors. By removing these intensity-dependent errors QM can therefore potentially lead to an improved climate change signal. The manuscript is very interesting and usually well written and deserves publication after some minor corrections:“

Thanks for your constructive comments and suggestions. We added our responses below your original comments:

“1) Figure 4: I found the colors used for different lines very confusing (eg Feb, March, Aug and September are difficult to differentiate). I would prefer to group seasons according to a similar color schemes (e.g blue for winter, green for spring, etc.) Also, it is striking to me how the model error characteristic in, e.g., IP (SMHI) changes so drastically from Jan to Feb, passing from a positive to a negative slope. Is there any plausible explanation for that behavior?”

Thanks for the suggestion. We changed the colors in Fig. 4 in a way that each month in a season has the same color (e.g., the winter months December, January and February are now blue). Within the seasons we discriminate between individual months by using different line styles. The new figure 4 is attached to for illustration. The same modifications have been applied to all similar figures in the supplementary material (Figs. S1 – S8) and the captions have been adapted accordingly.
Regarding the explanation of the difference between the Jan and the Feb error characteristics of the SMHI model in IP, we refrain from guessing. Such interpretation is outside the scope of the study. We don’t aim to analyse the errors of each individual model in each region in detail, since the focus of this study is not on model development, nor the physical explanation of model errors. We take them as given and focus on their influence on the climate change signal and on empirical-statistical post-processing of the model results.



Modified Fig. 4 of the manuscript using an enhanced colour scheme, as suggested by the reviewer.

“2) Figure 5: is the bold line (“ensemble average error characteristic”) the ensemble mean of the individual models’ errors, or the error of the multi-model mean? As in many works it is claimed that the MM mean usually outperforms any single models, would it be possible to show the error characteristic for the MM mean as well?”

We show the average of the individual model’s daily temperature error characteristics, not the error characteristics of the daily ensemble mean temperature. Both would be identical for the bias, but not necessarily for the “error characteristics” (i.e. the ECDF). We show the former, since it directly relates to the analysis we perform and to the quantities described in the formulae of our study. Although the question raised by the referee (does the multi-model mean outperform individual models?) is interesting, it is not directly relevant for our analysis. Therefore, and since we don’t want to distract from the main topic of the study, we prefer not to add the error of the multi-model mean temperature to Fig. 5.

“3) Is there any reference for Eq 2?”

We are not aware of any reference for Eq. 2. This simple formula has been originally designed for this study and represents a simplified (linearized) model of “intensity-dependent” model errors. Its explanation is given in the paper. In particular, please refer to section 2.1 for the discussion of the concept of intensity-dependence and to the paragraph directly after Eq. 2 for a brief discussion of the validity of the linearization.

“4) I have some problems with the notation of eq 6 (and similarly, eq 8) Is Delta Y supposed to be Delta Yi ? And similarly is cov(s,Dy) supposed to be cov(si,Dy’i)? If the authors chose to change the notation for clarity, they should specify it in the text. Unless I am wrong about the notation, but then I do not understand eq 6, as Dy is not defined in the text, for instance.”

Since our current notation of equation 6 and the following equations is misleading, we changed all affected equations following the suggestion of the referee.

“5) In fig 8 it is striking how QM and LC give sometimes opposite results. The authors briefly address this point claiming that it needs further analysis. In my opinion, the fact that the QM method applied here uses the same constant correction outside the calibration range is a major point. Would it be possible to perform a simple test (on only one month for only one model) by using a QM method with a linear correction even outside the calibration range and to compare it with both the original QM and the LC?”

The differences between QM and LC pointed out by the Referee are unquestionably disturbing. However, we didn't find a way to clearly identify the reason within the time-resources we had for this study. A set of experiments with clearly defined data- and model error characteristics would be needed to distinguish between different potential reasons (i.e. this would involve carefully designed artificial data and artificial models). This was beyond the resources we had for this study and we hope it is acceptable for the referee to leave it to further research.